

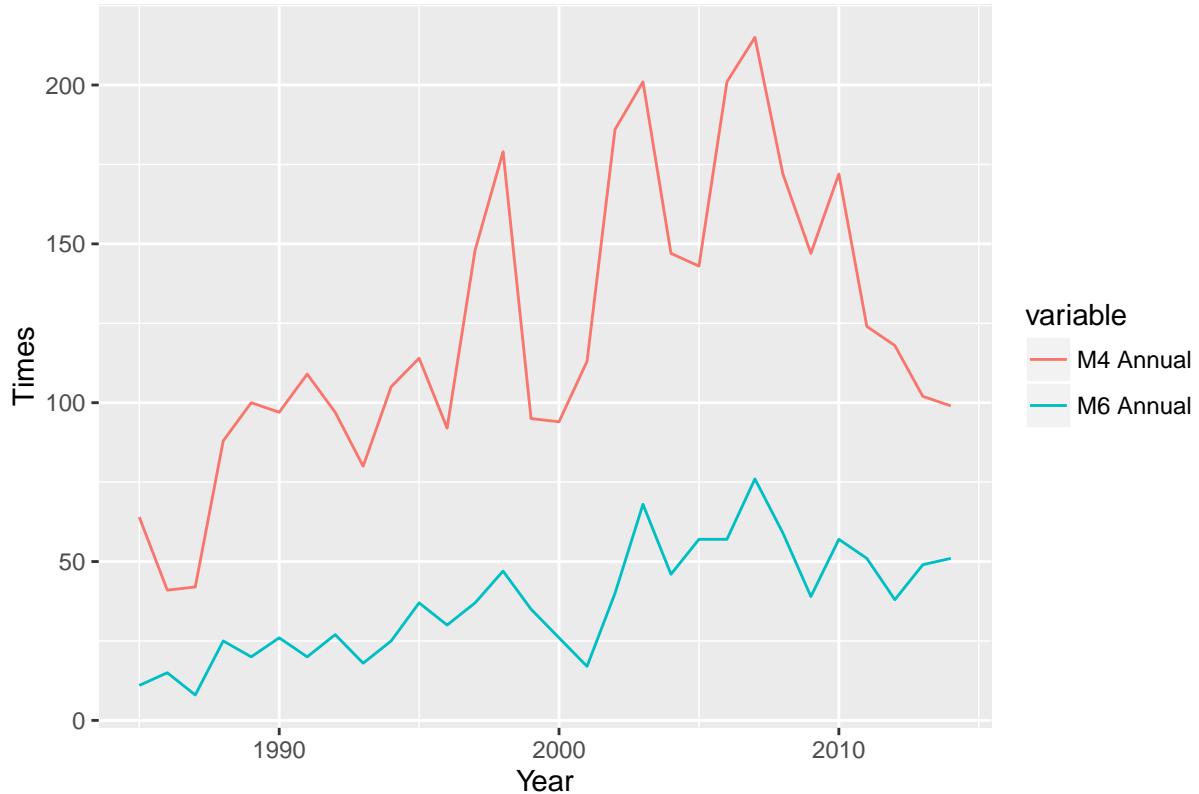
report2_final

ggMonet

March 8, 2016

Time Series Plot of Floods

Annual Flood Time Series Plot



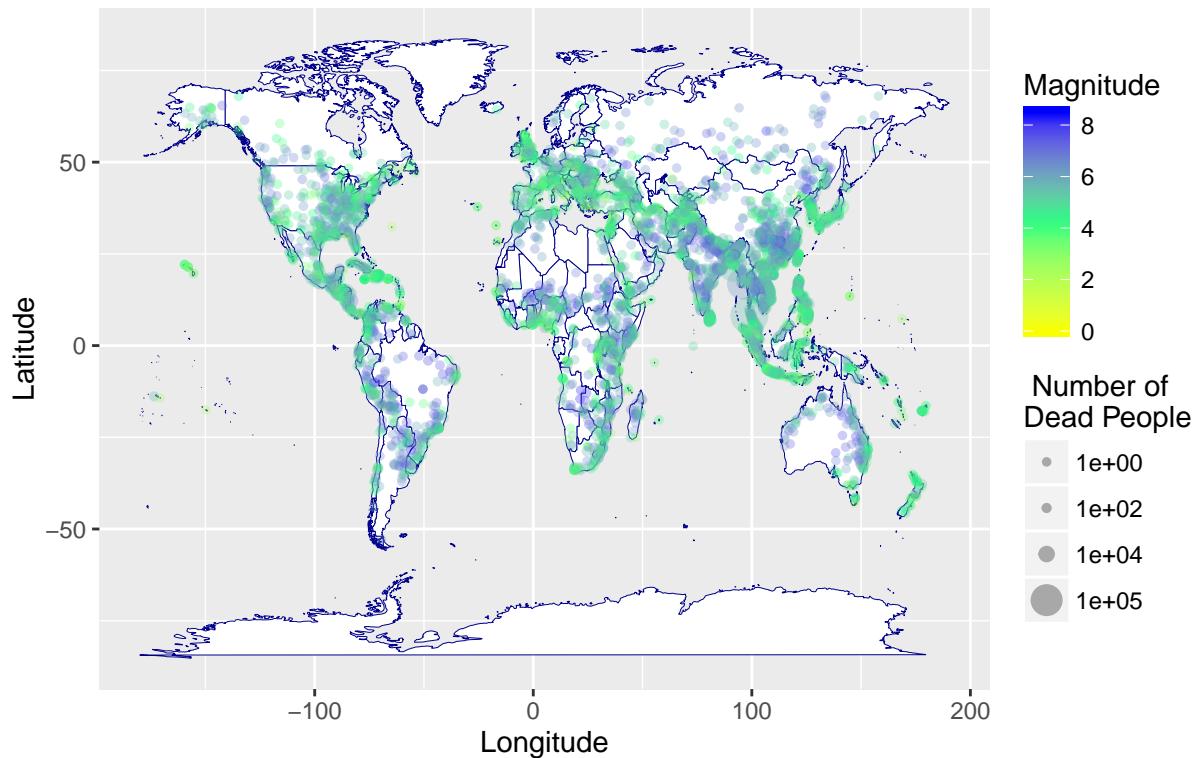
We can see a greater fluctuation in the number of M4 floods than M6 floods, and the occurrences of floods of both degrees are highly correlated. After a view of the distribution of floods in time, we will present the spatial distribution.

Spatial Distribution of Floods

```
## [1] "numeric"
```

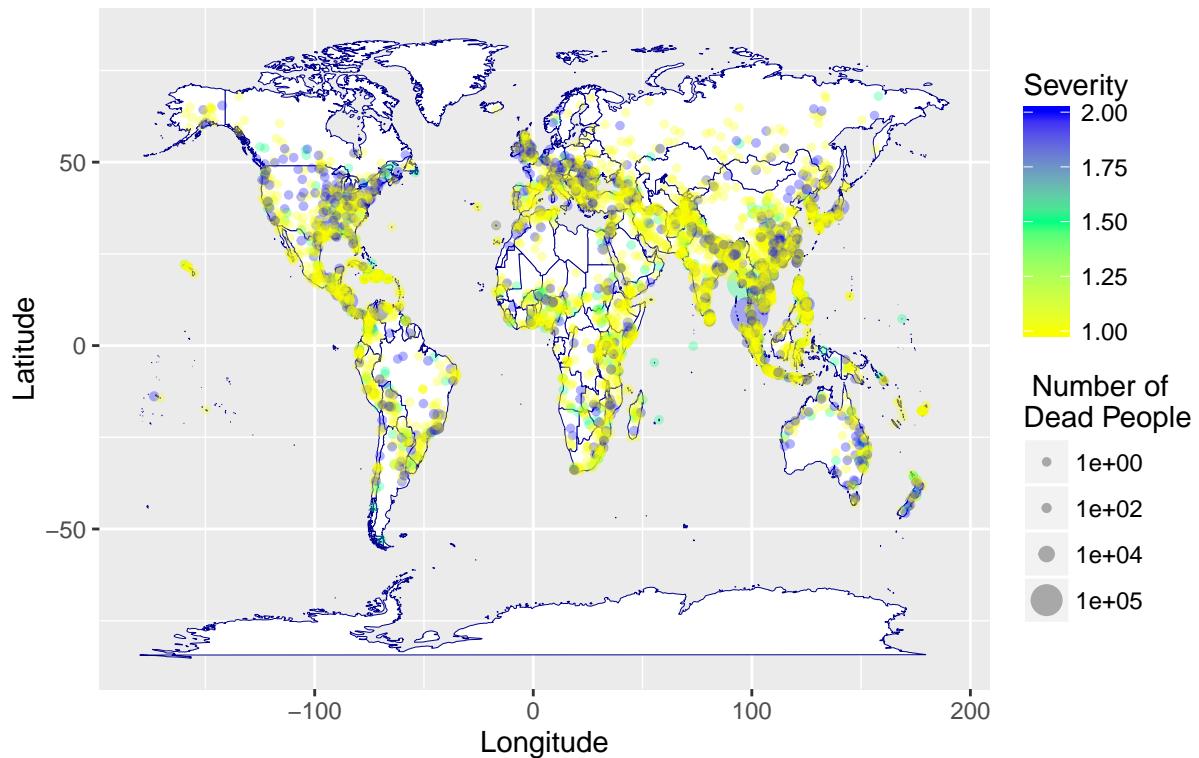
We preprocessed data by identifying the top 4 main reasons among these 10 causes based on some key words.

Flood Distribution with Number of Dead People and Magnitude

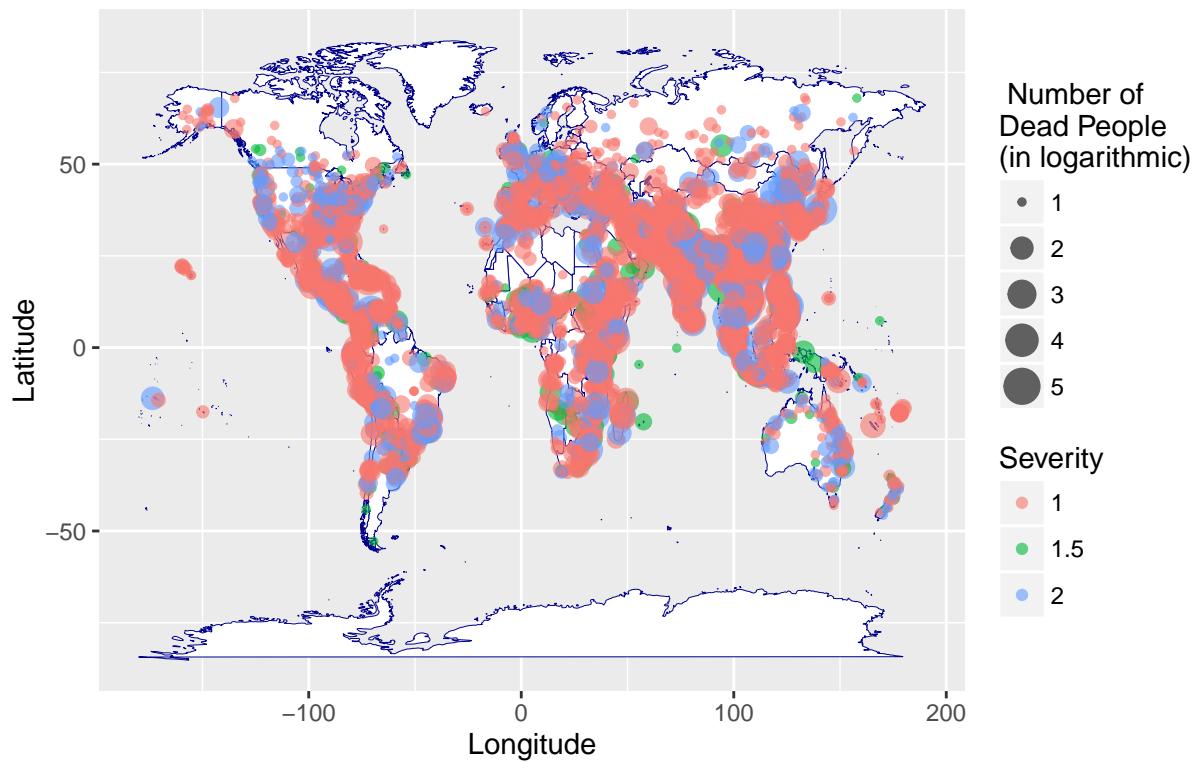


From this distribution plot, combining the magnitude and the number of deaths, we can learn that the floods in East Asia and South Asia had both higher magnitude and larger number of deaths. Specifically, in Malaysia and Thailand, there were several highly serious floods, which resulted in over one hundred thousand people losing their lives. Besides, the East US and Europe also had comparatively more floods with lower magnitude. However, in the southeast coast of South America, the floods suffered from higher magnitudes on average.

Flood Distribution with Number of Dead People and Severity



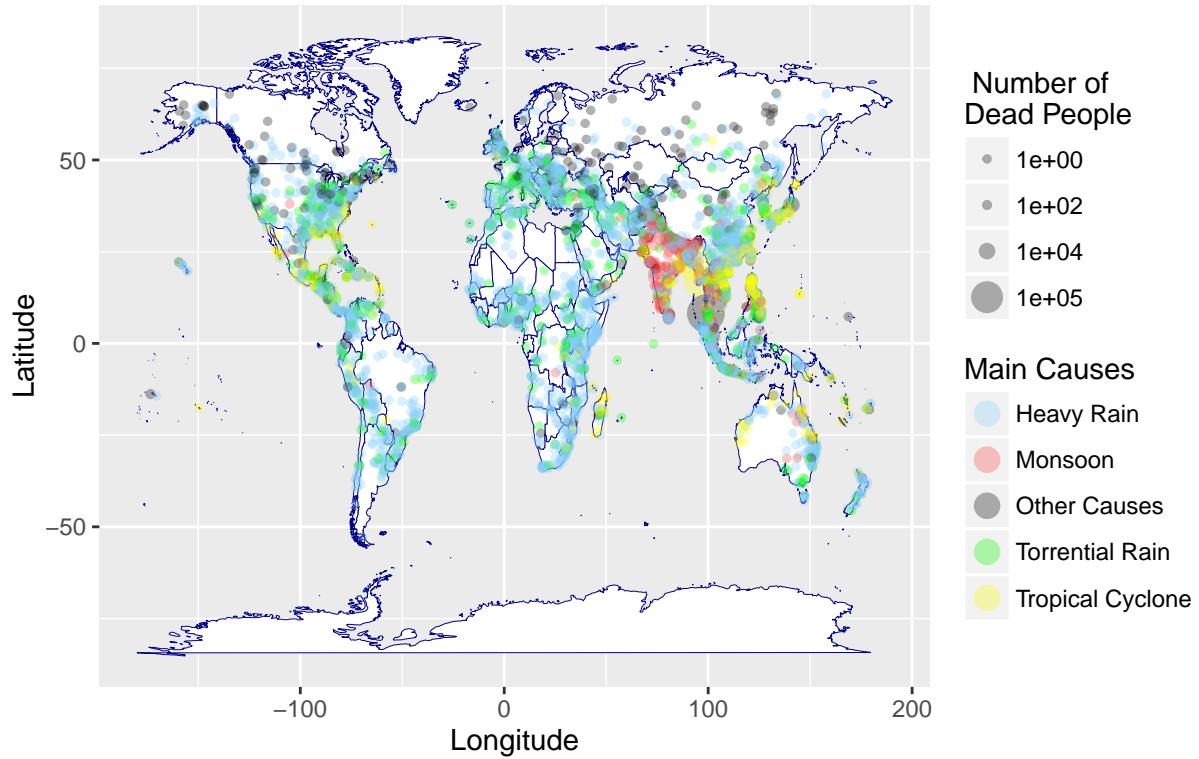
Flood Distribution with Number of Dead People and Severity



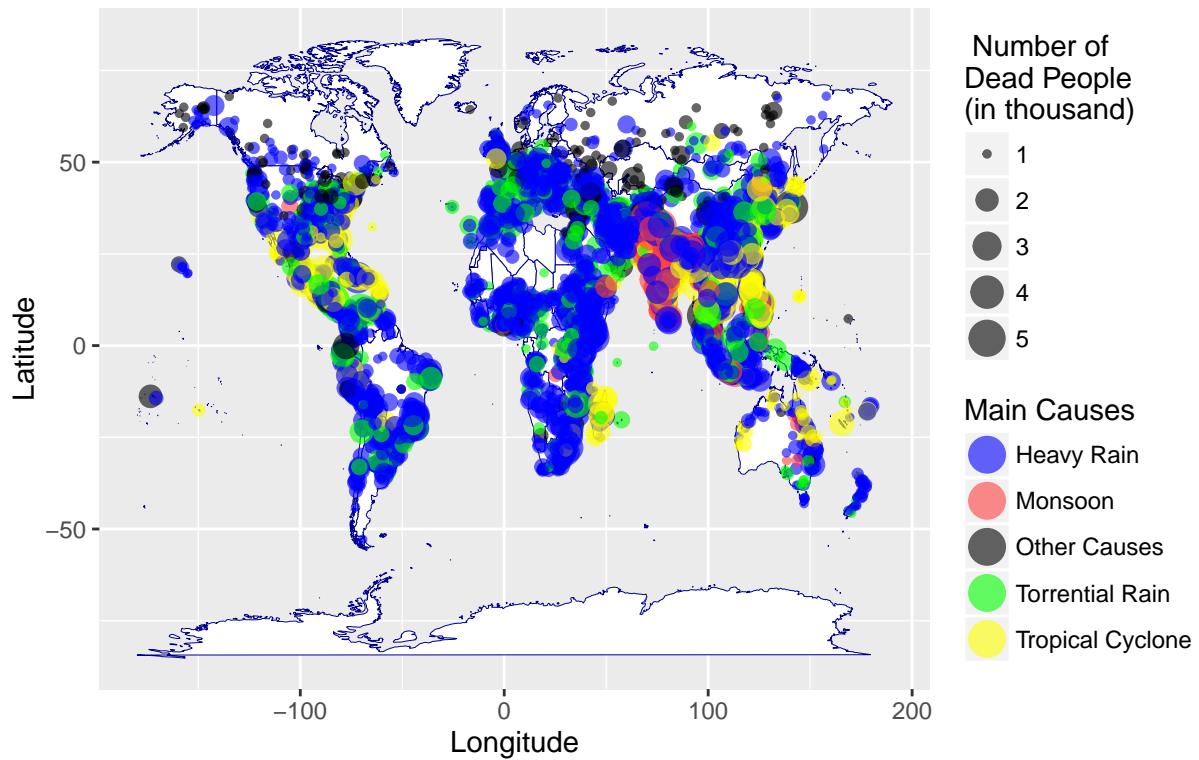
From this distribution plot, combining the severity and the number of deaths, we can see that Europe had

the densest floods with high severity. Compared to these, the floods in Asia had higher densities but lower severity. But still, the most serious flood around Malaysia had the most severity and the largest number of deaths at the same time. We also notice that in other places like Russia and Africa, the floods had the lowest severities.

Flood Distribution with Number of Dead People and Main Causes

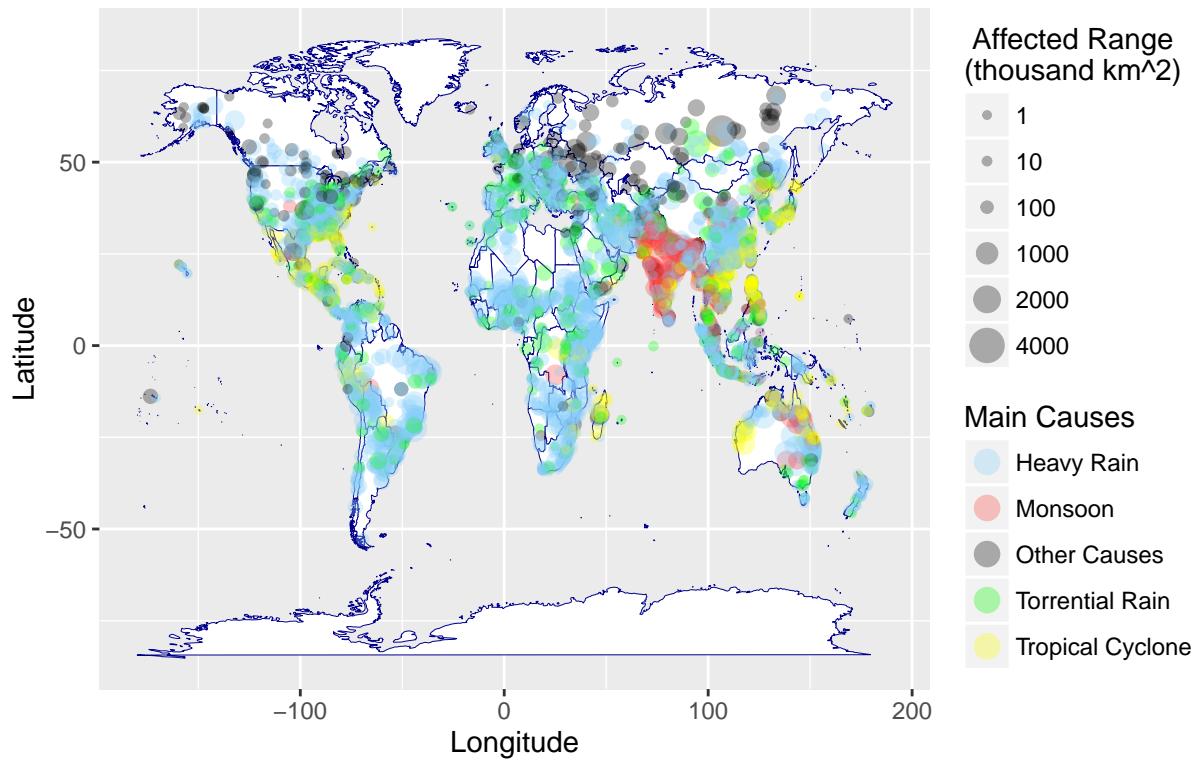


Flood Distribution with Number of Dead People and Main Causes



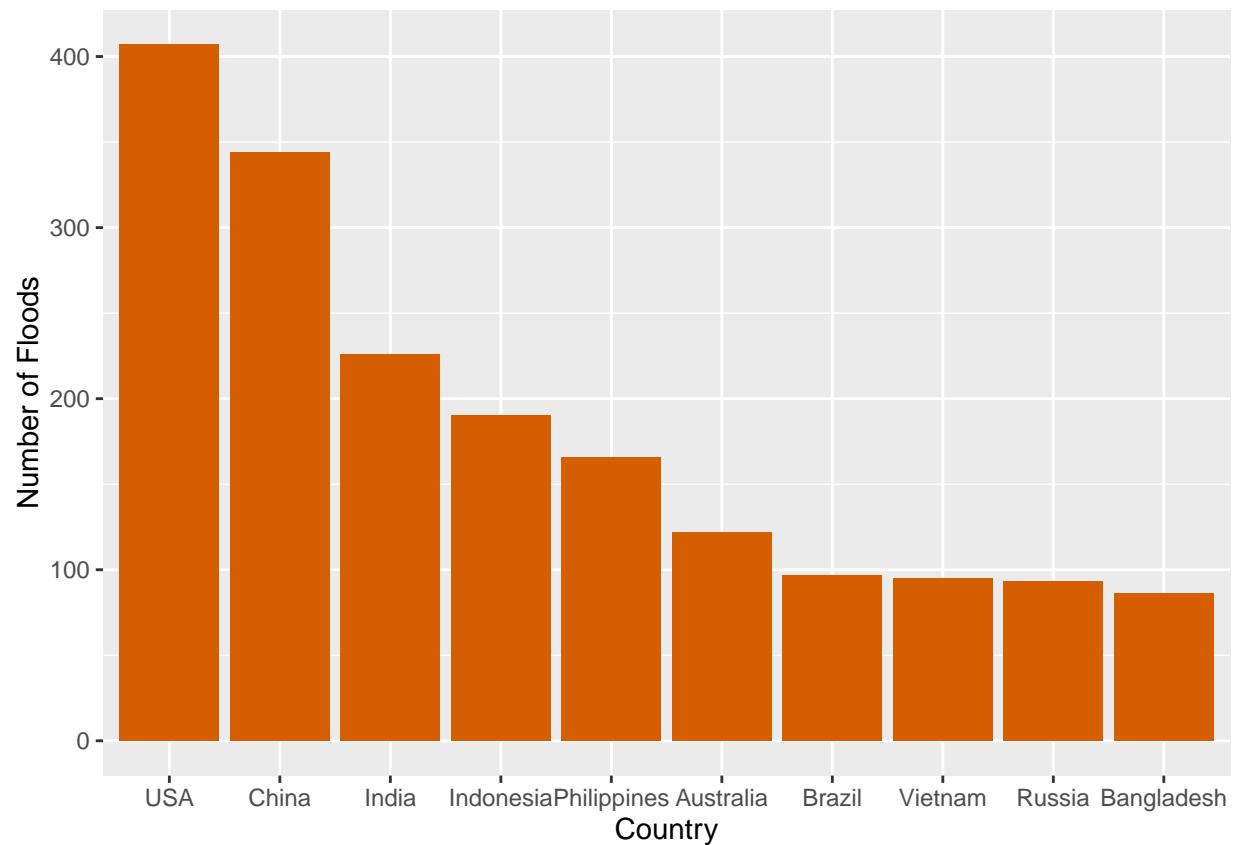
This time, we analyze the relationship between the main reasons behind these floods and the number of deaths. From the corresponding distribution plot, the following conclusions can be drawn. Firstly, the most common reason in the whole wide world is Heavy Rain, especially in Africa. Then, we find the floods in India mostly resulted from Monsoon, which seems to be the exclusive reason. Besides, the floods cased by tropical Cyclone were concentrated on the South Asia and South US. The dense floods in Europe were caused by a combination of Heavy Rain, Torrential Rain and other reasons. As for the floods in Russia, the reasons behind them seems to be uncommon, at least compared to other places.

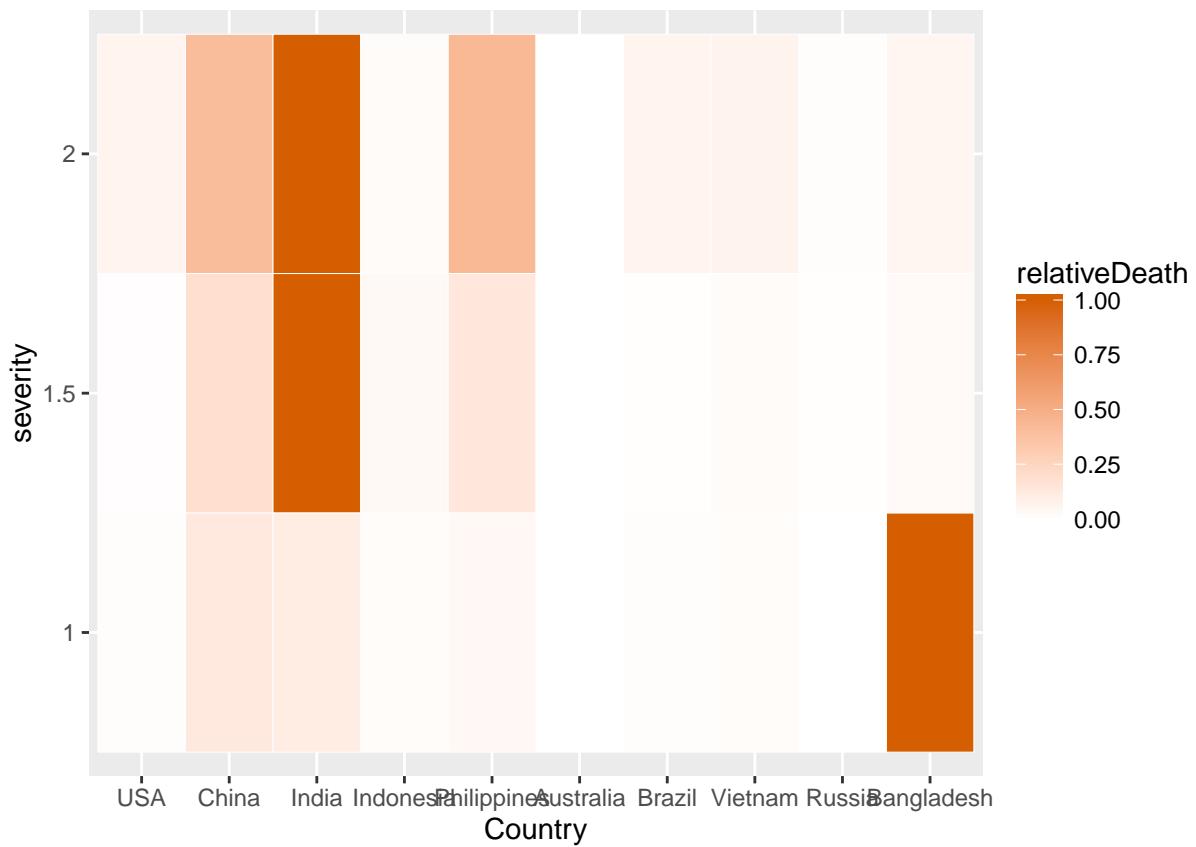
Flood Distribution Based on Main Causes and Affected Range(km^2)



In the end, we compare the distributions between the main causes and the affected range (square kilometer). From the picture, we can see that the floods in India caused by Monsoon seem to have the one of largest affected scopes. Besides, the causes behinds the intensive floods in South Asia were highly complicated, leading to comparatively small affected ranges with surprise. For example, the most terrible flood in Malaysia that we've discussed seems to have an unexpected small range. The situations of East China and East US seem to be similar: large affected ranges and various reasons behind these dense floods. On the top of these coast areas, the floods in Russia mainland also had large affected ranges.

Floods in Several Countries

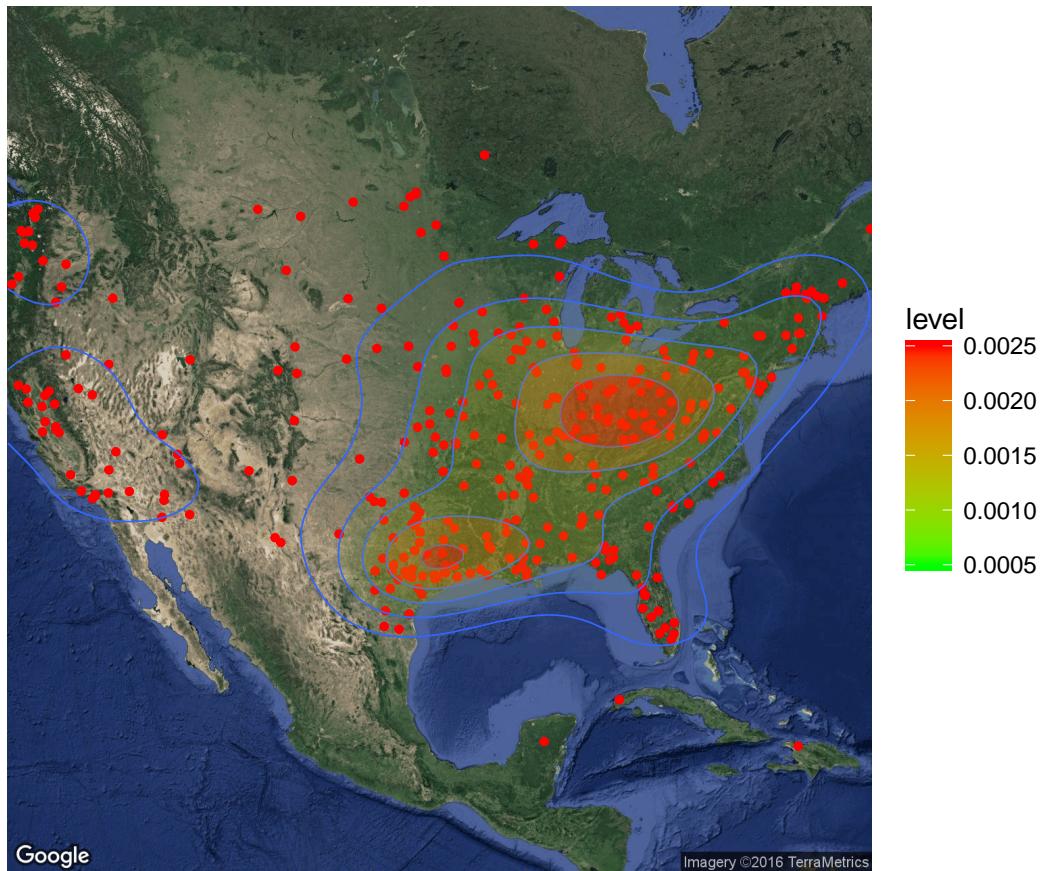




As we can see, USA, China and India are the top three frequently impacted countries. In the next part, we will have a more detailed look into the distributions of floods happened in the three countries. And from the heatmap of the relative death (rescaled by the most death occurred in a certain severity), we can see that USA did a great job in preventing death in floods while India and Bangladesh did not.

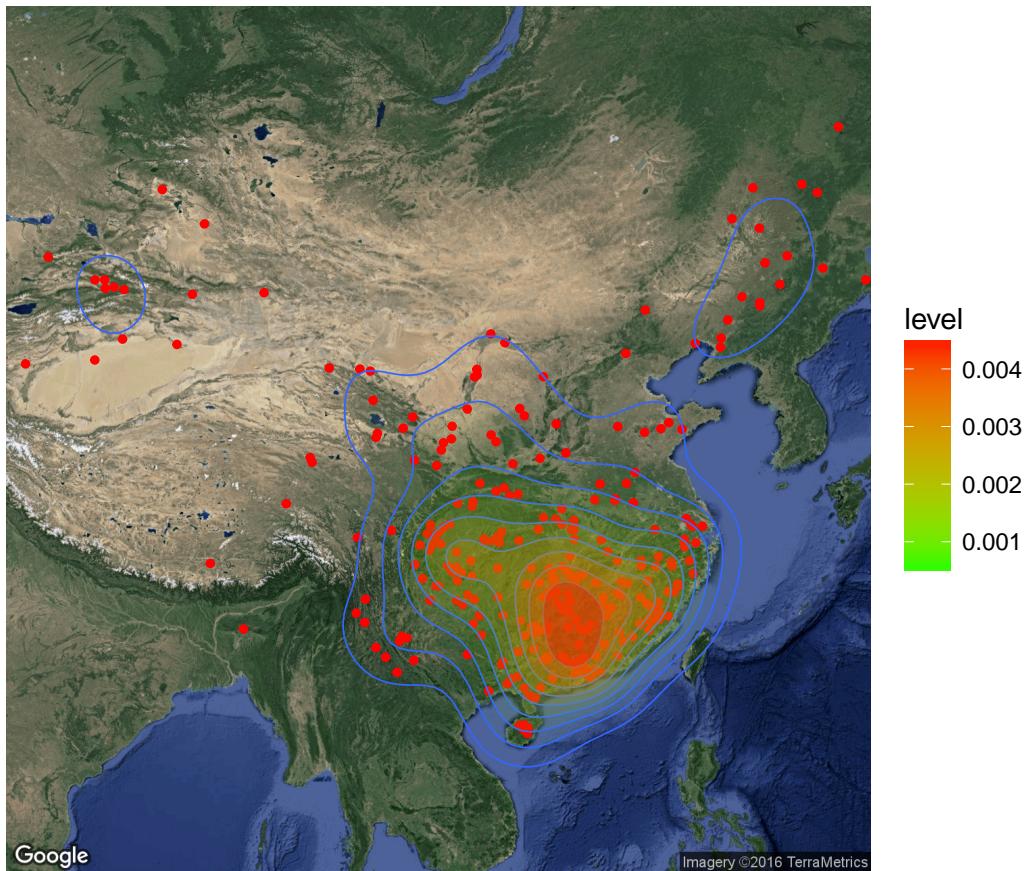
Here, we will describe distribution of floods in USA, China and India.

Firstly, we will draw a case of USA.



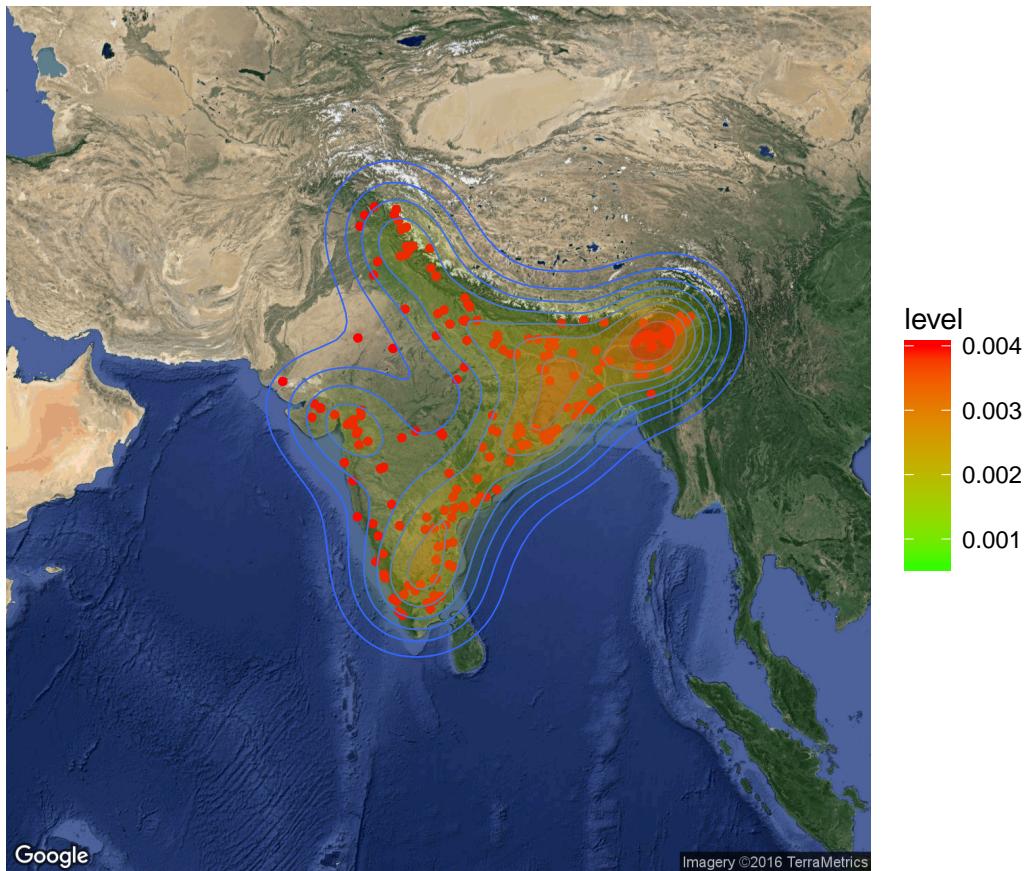
We draw contours based on where floods happened in the past. At the areas which have red color, they had floods many times. As the color turns green, the numbers of floods occurrence decrease. In the case of USA, floods often happened at coastlands. Also, we can see that floods often happened at the middle east area.

Secondly, we will draw a case of China.



In the same way as USA, floods often happened at coastlands in China. We can see that at the area of south east, floods frequently happened.

Thirdly, we will draw a case of India.



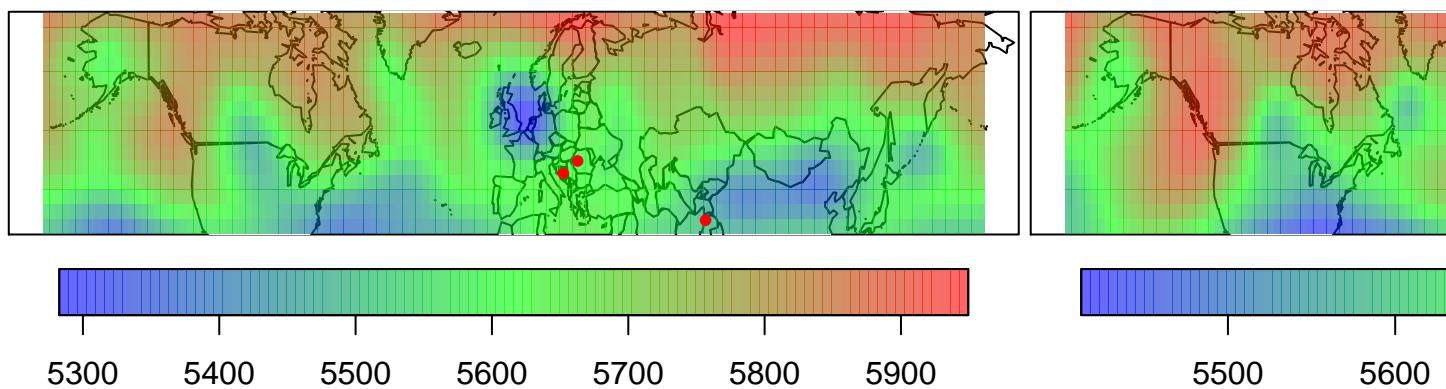
We can see that floods often happened at coastlands in India, and at east area of India, floods frequently happened as well.

Pressure Distribution of Days With More than 4 Floods

We selected several days with the most occurrences of floods and plotted the distribution of the pressure of given area.

```
## [1] "2010-06-22" "2010-07-27"
```

2010-06-22

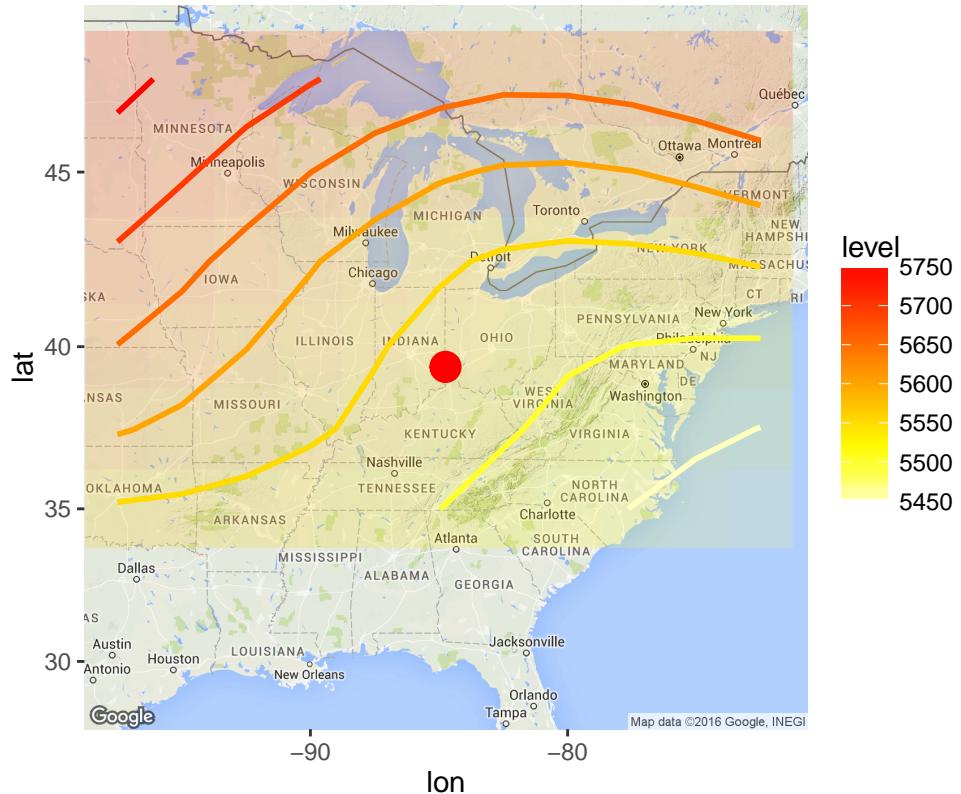


According to the plots, we can see that many floods occurred in the areas with lower pressure, detailed analysis will be illustrated below.

Pressure Distribution of Several Floods

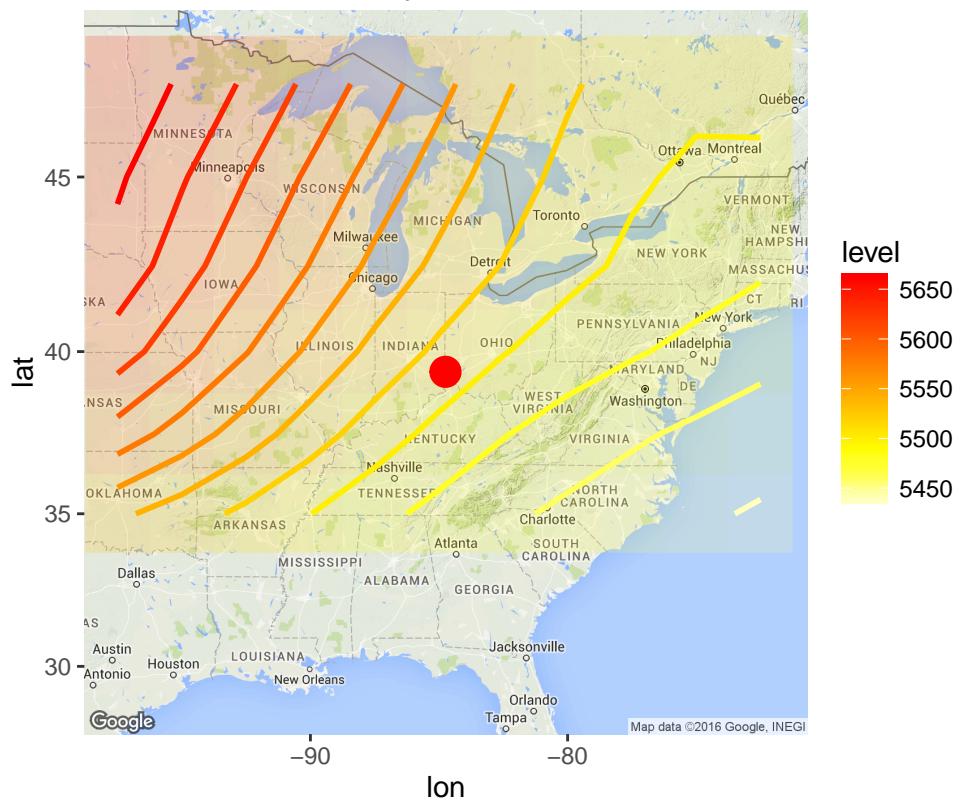
Firstly, we will focus on changes of contours during the flood that happened in the area of Southern Michigan, central Indiana, and western Ohio in the US from Jun 27, 2015 to Jun 29, 2015.

5 days before the flood happened



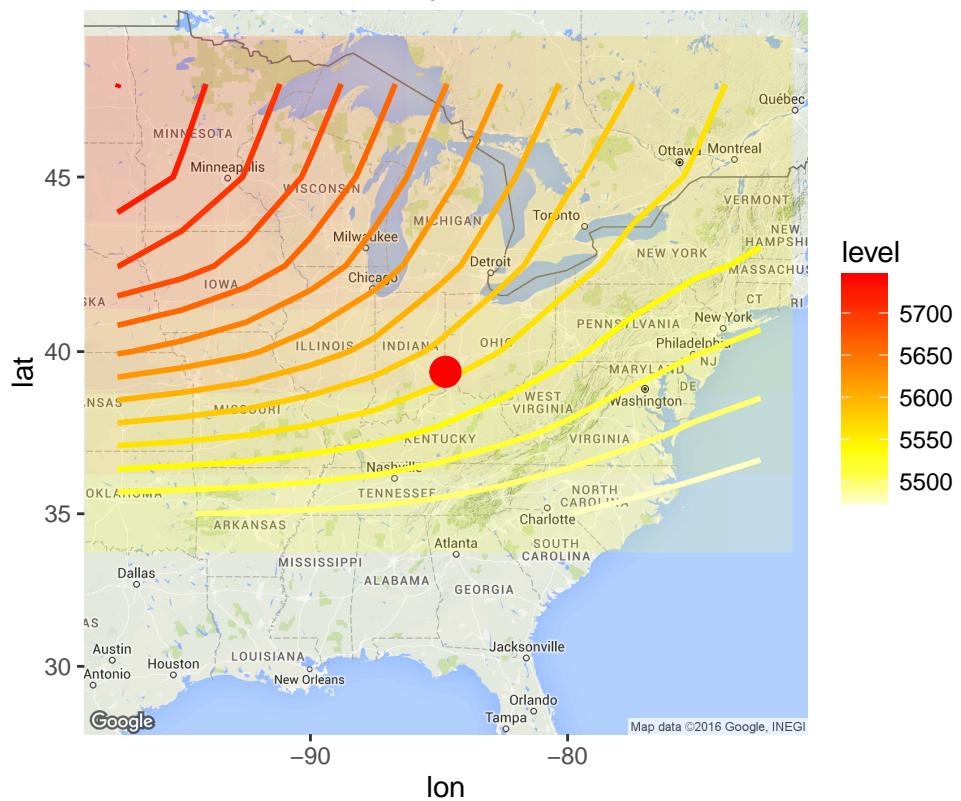
This is a contour information of pressures at Jun 22, 2015. We also draw where flood happened by using red dot. At “level” bar on the right of the map, values of contour are written. For example, the area that has deep red has high pressure. We can see that pressures between each area is wide.

First day of the flood



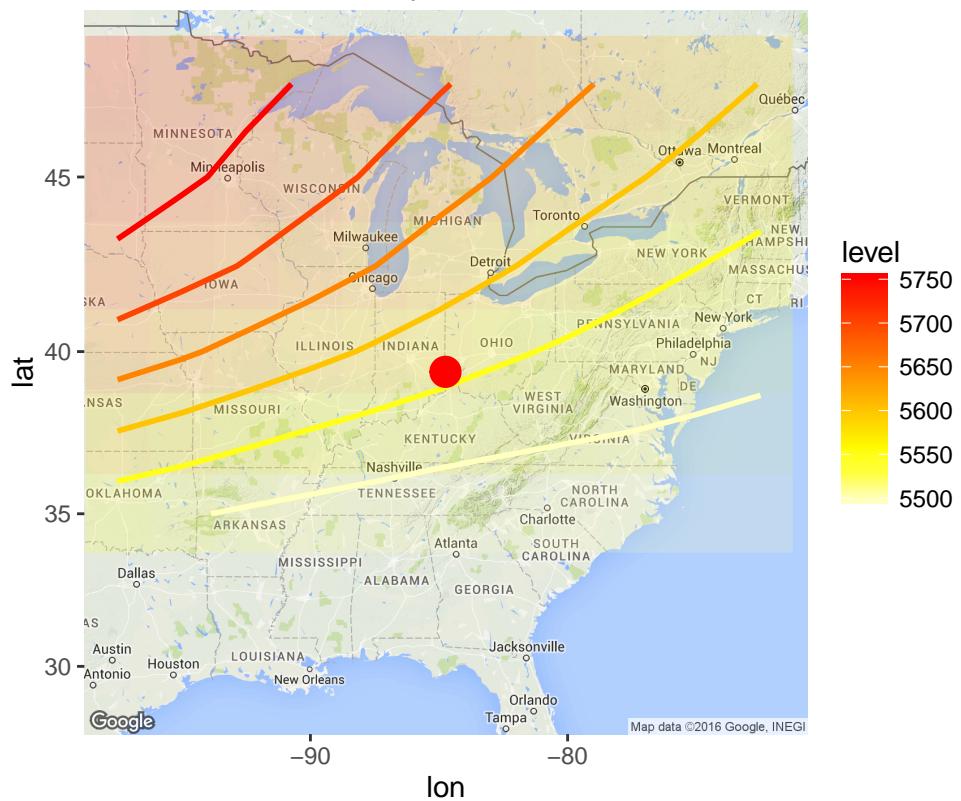
This is a contour information of pressures at Jun 27, 2015. As we can see, pressures between each area is narrow, and there are high range of pressures in this area. Also, at the flood point, they had a value of the pressure that was around 5500.

Second day of the flood



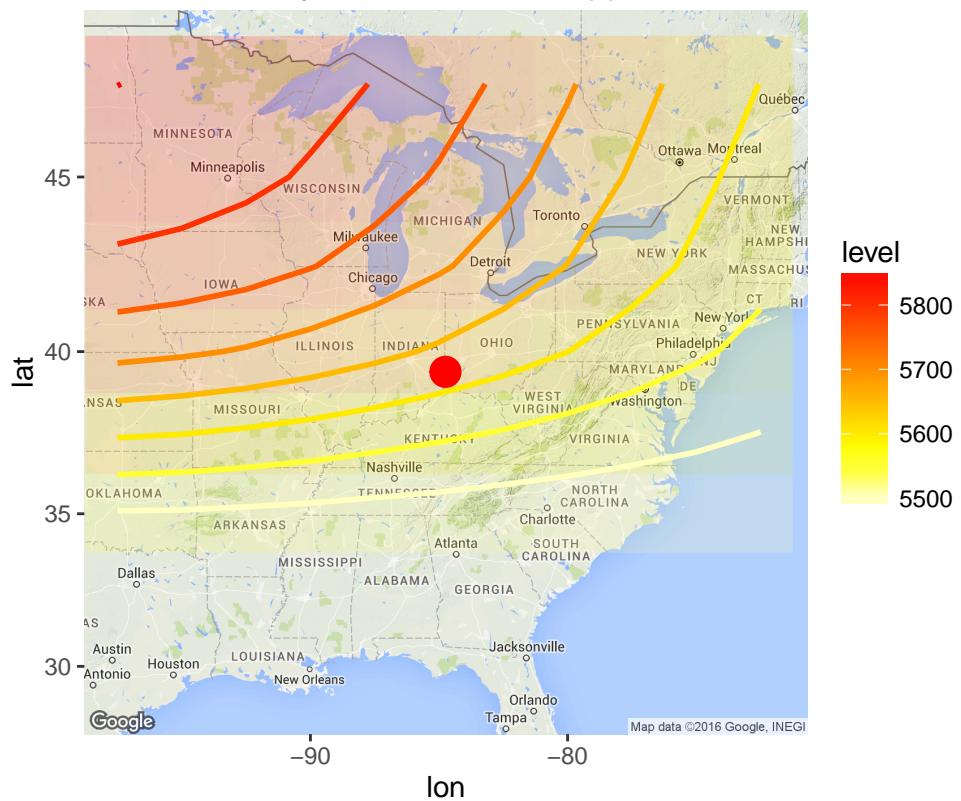
This is a contour information of pressures at Jun 28, 2015. As we can see, pressures between each area is narrow as well. At the flood point, they had a value of the pressure that was around 5500.

Third day of the flood



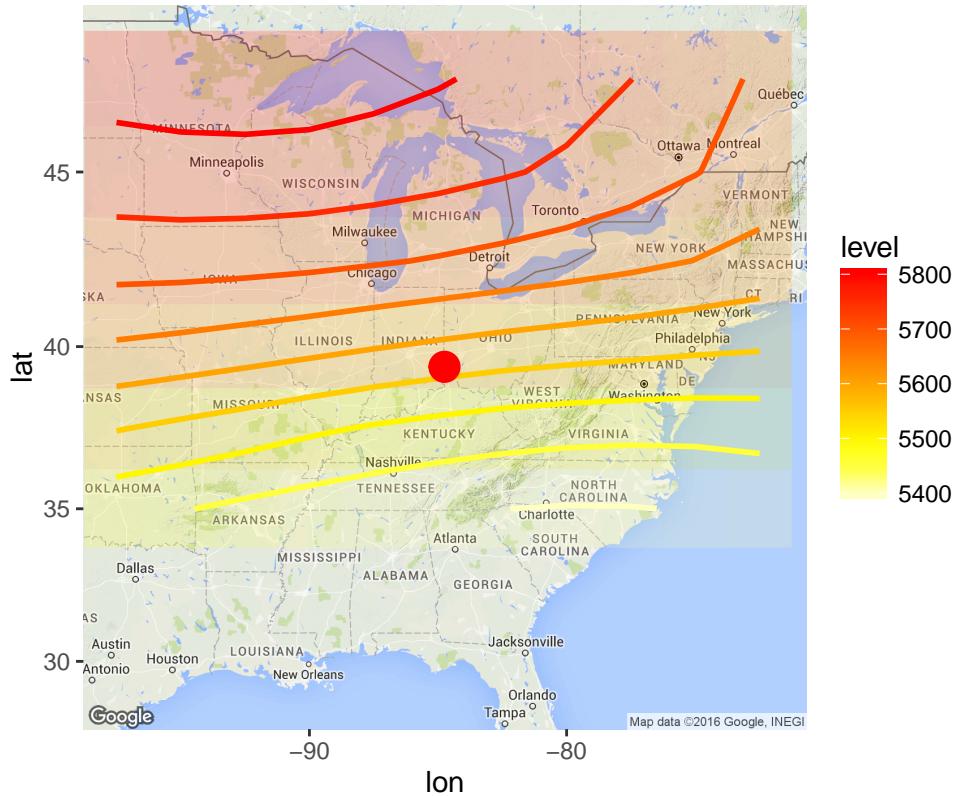
This is a contour information of pressures at Jun 29, 2015. Pressures between each area becomes wider. At the flood point, they had a value of the pressure that was around 5550.

One day after the flood happened



This is a contour information of pressures at Jun 30, 2015. It seems that they had higher pressure value (around 5600) compared with data while flood was happenin.

Two days after the flood happened



This is a contour information of pressures at July 1, 2015. As same with the previous map, they had higher pressure value (around 5600) compared with data while flood was happenin.

PCA of Pressure Data

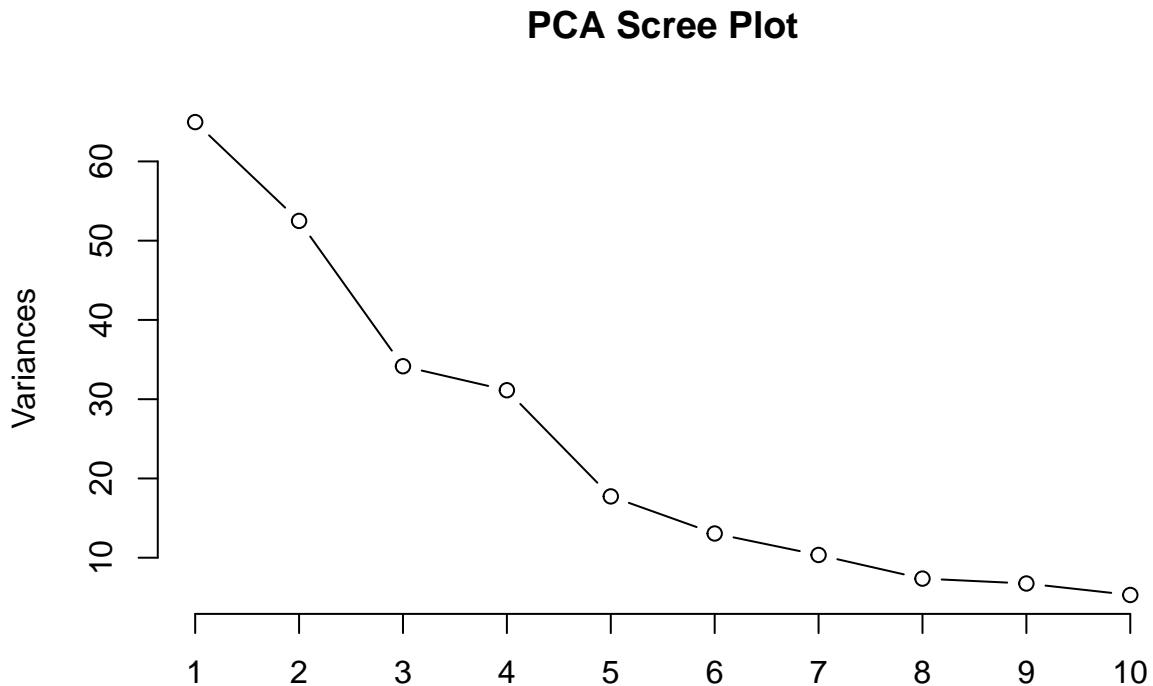
In performing PCA on the NOAA data, the pre-processing involved re-formatting the data into an $N \times D$ matrix, such that: $N = \text{days}$, and $D = \text{lon} \times \text{lat}$ coordinates. Specifically, our goal was to do PCA over the USA, during the months of June and July, 2015 (to examine the floods covered in the above section).

In this case, focusing on a specific location and timeframe was the only way to make PCA both intelligible and computationally feasible. Per the NOAA protocol, time is defined in days since 1/1/1948, so in this case, analyzing June-July 2015 corresponds with days 24624 to 24683. With regards to geography, the longitude and latitude will be bounded by: 230-300 Degrees East, and 25-55 Degrees North, respectively.

Doing the PCA itself involved centering and scaling each “grid” of longXlat data (in order to standardize the PCA). Here is the summary output of the PCA, focusing on the first 10 Principal Components:

```
## Importance of components:
##                               PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation     8.060  7.2453  5.8443  5.5795  4.2113  3.61319  3.21658
## Proportion of Variance 0.267  0.2158  0.1404  0.1280  0.0729  0.05366  0.04253
## Cumulative Proportion  0.267  0.4828  0.6232  0.7511  0.8240  0.87769  0.92022
##                               PC8      PC9      PC10
## Standard deviation     2.71364 2.59776 2.30177
## Proportion of Variance 0.03027 0.02774 0.02178
## Cumulative Proportion  0.95048 0.97822 1.00000
```

This is a Scree Plot showing the relative importance of the first 10 PCs.



Finally, we used the `ggbio` package to build an interesting, though dense, biplot of the PCA data.

```
## Skipping install for github remote, the SHA1 (7325e880) has not changed since last install.  
## Use `force = TRUE` to force installation
```

